# Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis

Kevin J. Boudreau
London Business School, London NW1 4SA, United Kingdom, kboudreau@london.edu

Nicola Lacetera
Rotman School of Management, University of Toronto, Toronto, Ontario M5S 2E9, Canada,
nicola.lacetera@utoronto.ca

Karim R. Lakhani
Harvard Business School, Boston, Massachusetts 02163, k@hbs.edu

Contests are a historically important and increasingly popular mechanism for encouraging innovation. A central concern in designing innovation contests is how many competitors to admit. Using a unique data set of 9,661 software contests, we provide evidence of two coexisting and opposing forces that operate when the number of competitors increases. Greater rivalry reduces the incentives of all competitors in a contest to exert effort and make investments. At the same time, adding competitors increases the likelihood that at least one competitor will find an extreme-value solution. We show that the effort-reducing effect of greater rivalry dominates for less uncertain problems, whereas the effect on the extreme value prevails for more uncertain problems. Adding competitors thus systematically increases overall contest performance for high-uncertainty problems. We also find that higher uncertainty reduces the negative effect of added competitors on incentives. Thus, uncertainty and the nature of the problem should be explicitly considered in the design of innovation tournaments. We explore the implications of our findings for the theory and practice of innovation contests.

*Key words*: innovation contests; uncertainty; innovation; problem solving; tournaments
*History*: Received July 26, 2009; accepted December 22, 2010, by Christian Terwiesch, operations management. Published online in *Articles in Advance*.

## 1. Introduction

Contests are a well-established mechanism for eliciting innovation (Terwiesch and Ulrich 2009, Terwiesch and Xu 2008, Scotchmer 2004), and calls for their use are increasingly frequent in the private and public sectors.[1] It is currently estimated that the "contests industry" might have a value between $1 and $2 billion (McKinsey & Company 2009). A long-standing question within the literature and practice has been "How 'big' should an innovation contest be?" or "How many competitors should be admitted to a contest?" (Che and Gale 2003, Fullerton and McAfee 1999, Taylor 1995, Terwiesch and Xu 2008). Research in economics suggests that increasing the number of competitors who are admitted to a contest will reduce the likelihood of any one competitor winning, thereby reducing incentives to invest or exert effort and lowering overall innovation outcomes (Che and Gale 2003,

Fullerton and McAfee 1999, Taylor 1995).[2] Similar predictions and findings on negative incentive effects have been found in research in sociology and psychology (Bothner et al. 2007, Garcia and Tor 2009). Overall, the literature has generally recommended against free entry into contests, with some models specifically determining the ideal number of competitors to be just two (Che and Gale 2003, Fullerton and McAfee 1999).

Although there are cases of contest sponsors deliberately restricting the size of their contests (McKinsey & Company 2009, Nasar 1999), historical and modern examples of innovation contests include many cases in which sponsors encourage large numbers of competitors to enter. In the 15th century, the office responsible for the construction of Florence's new cathedral, Santa Maria del Fiore, announced in 1418 a contest to solve a 50-year-old architectural

---

[1] See, for example, Lindegaard (2010), McKinsey & Company (2009), National Research Council (2007), Tapscott and Williams (2006), and White House (2010).

[2] Beyond preserving incentives, another reason to limit entry in a contest is to decrease the cost to the contest organizer conducting and evaluating the competition (Fullerton and McAfee 1999).

puzzle—the creation of the world's widest and tallest dome—with an open invitation for anyone to participate. The organizers received more than a dozen design proposals and deliberated for more than a year before selecting one from an unexpected source, goldsmith and clockmaker Filippo Brunelleschi (King 2000). More recently, in 2000, the Canadian mining company Goldcorp announced a $500,000 contest aimed at discovering new gold targets in a low-performing Northern Ontario mine. It, too, encouraged widespread entry; the contest attracted more than 1,400 participants and led to the remarkable discovery of 44 new, productive targets (Tapscott and Williams 2006). In 2006, the 2006 Netflix Prize contest, established to develop software that would achieve a 10% improvement in the DVD rental firm's algorithm-based movie recommendation system and including a prize of $1 M, received submissions from 5,169 individuals and teams.[3] Apart from these ad hoc contests, firms have begun to set up contest platforms as an ongoing business model (Boudreau and Hagiu 2009). InnoCentive.com, for example, routinely attracts roughly 250 individuals to contests involving R&D-related scientific problem solving on behalf of its clients (Jeppesen and Lakhani 2010). Thus, rather than restrict entry, the tendency has been to open innovation contests to all comers. This would appear to contradict mainstream economic theory yet has remained prevalent in practice. Why is this so?

One possible explanation is that the quality of any one solution, including the solution developed by the eventual winner, does not just depend on how much effort is exerted or even on a competitor's skills or aptitude. There may remain substantial uncertainty regarding how best to approach and solve an innovation problem. The problem may require a novel solution, one that has yet to be discovered. Precisely who will win and the best technical approach may be hard to anticipate. Having a large number of competitors in an innovation contest may simply increase the likelihood of finding at least one particularly good solution—in other words, an extreme-value outcome. This perspective is consistent with the literature, which highlights that increasing the number of independent experiments—pursuing independent approaches or "parallel paths" along the technical frontier—can improve overall innovative performance (Abernathy and Rosenbloom 1969, Dahan and Mendelson 2001, Nelson 1961). Hence, we refer to the possibility that adding greater numbers of competitors will lead to a greater chance of extreme outcomes as the "parallel path effect." This might be particularly important where innovation managers in general and contest organizers in particular care about

the maximum or best innovation performance above anything else (Dahan and Mendelson 2001, Girotra et al. 2010, Terwiesch and Ulrich 2009).

Analyses of parallel paths and incentive effects proceeded in largely independent literatures until Terwiesch and Xu (2008) proposed an approach to integrate the two mechanisms within the same analytical framework. This required merging the order-statistic modeling apparatus of parallel path models with systematic modeling of strategic interactions and incentives. Terwiesch and Xu argued that adding greater numbers of competitors generates a tension between the negative effects on incentives and the positive effects of parallel paths, leading to particular instances in which free entry or limited entry would generate better outcomes depending on the particular parameters in their model. The analysis also highlighted the importance of the maximum or winning score in a contest. If such a tension were to exist and were to be empirically relevant, the optimal size of a contest should be larger than an analysis of economic incentives would suggest.

Given the potential importance of these effects on the central question of how big a contest should be, our goal is to test for the existence of a trade-off and the interplay between incentives and parallel path effects, thus providing an empirical foundation for the recent theoretical advances. We address three related questions: (1) Are incentive and parallel path effects of comparable magnitude and, consequently, do they need to be explicitly considered together when designing contests? (2) Do the incentive and parallel path effects work as simply as has been theorized, one effect dampening, the other stimulating innovation? (3) Under what conditions might one effect dominate the other?

Addressing these questions empirically is challenging. A most basic condition is simply finding a large sample of comparable innovation contests, given that contests often exist as one-off endeavors by their very nature. To discern the particular mechanisms at work, we also require precise measures of key microeconomic variables, including objective measures of innovation outcomes. Apart from observing multiple trials, we require some source of random (exogenous) variation across these trials, particularly in the number of competitors. We also have the requirement here of not only observing contest winners but also the entire distribution of outcomes.

In this paper, we have a unique opportunity to use a context and data that satisfy these requirements in a natural setting. We analyze 9,661 competitions related to the solution of 645 problems from Top-Coder, a contest platform where elite software developers participate in regularly held competitions to create novel software algorithms. This environment

---

[3] Data obtained from http://www.netflixprize.com//leaderboard.

affords the opportunity to study multiple concurrent contests for the same problem with different numbers of direct competitors. Further, we are able to observe the skill level and quality of the solution for individual contestants.

Our analysis begins by estimating the independent workings of both incentive effects and parallel path effects. We confirm that these effects, when regarded separately, operate as is typically predicted in the theoretical literature. We also show, through quantile regressions, that the entire distribution of outcomes shifts downward with added competitors, as is usually predicted of incentive effects in one-shot innovation contests. In absolute terms, the shift is larger for outcomes in higher percentiles. However, when more competitors are added, the maximum score increases relative to the rest of the distribution of outcomes. Taken together, these findings demonstrate that adding competitors indeed worsens outcomes in expectation but increases the "upside" in that at least one competitor would achieve an extreme outcome. These effects are of comparable magnitudes in this context. Thus, neither of these effects can be ignored; both should be considered to assess the net effect of varying the size of a contest on problem-solving performance (i.e., the best performance within the group of competitors). These findings on their own should serve as a call for greater research into integrating and examining the interplay between parallel path and incentive effects in innovation contests.

We then highlight the key role played by the nature of the problem being solved in determining how big an innovation contest should be. We focus here on uncertainty about the sense of uncertainty in the best approach to solving a problem and, consequently, who will turn out to be the winner. In our context, uncertainty is closely related to the number of knowledge domains on which a problem solution draws. The basic idea here is that single domain problems are canonical problem types, with established solution approaches; multidomain problems are not simply additive but require novel out-of-paradigm solutions. We show that for problems drawing on a higher number of knowledge domains, or more uncertain problems, the parallel path effect is amplified, consistent with the higher likelihood of attaining an extreme outcome with higher uncertainty. We also find that higher uncertainty dampened the negative effect of added competitors on incentives in this context. Thus, more competitors could lead to improved contest performance but only when problems are highly uncertain and require a greater level of searching for the best approach or path to a solution.

Our findings suggest that considerable sensitivity to the relative importance of parallel path and incentive effects may be needed to design contests properly. On one hand, we might expect that the type of

problems that eventually are pushed out to contests might be characterized by considerable uncertainty and, thus, benefit more from large than from small, focused contests. On the other hand, the proliferation of contest platforms intended for repeated use might imply a preference for contests suited to a wider range of less uncertain problems for which a smaller number of competitors may be most desirable.

This paper proceeds as follows. Section 2 reviews the relevant literature and develops basic hypotheses. Section 3 details the empirical context of our study and describes the data and estimation strategies. The results of the empirical analyses are reported in §4. Section 5 summarizes our contribution and offers concluding remarks.

## 2. Literature and Hypothesis Development

This section reviews the literature on innovation contests, particularly as it centers on the effects of varying numbers of competitors. Our objective in this section is to develop three basic empirical hypotheses that will serve as a guiding set of predictions as we explore the nature of incentive and parallel path effects.

### 2.1. Contests and Incentives

Contests and relative performance evaluation mechanisms have received considerable attention in the economics literature, with examples drawn from political decision making, internal labor markets, sales performance contests, and sports (Casas-Arce and Martínez-Jerez 2009, Holmstrom 1982, Lazear and Rosen 1981). Research on innovation contests closely follows this tradition (Che and Gale 2003, Fullerton and McAfee 1999, Taylor 1995). A central question in this research is whether free entry or restricted numbers of participants should yield better outcomes.[4] The main intuitive message from existing models is as follows. In winner-takes-all contests with only one participant, contestants will have little incentive to exert effort to improve their work because there are no parties against whom they will be evaluated. Thus, adding some minimum level of competition should lead to greater effort (Harris and Vickers 1987). However,

---

[4] The other major issue in contest design addressed by the economics literature is how to set the "prices" (prizes, fees, and penalties) for contestants. Larger prizes tend to stimulate higher performance. Single prizes are argued to be effective for homogeneous and risk-neutral individuals; multiple prizes are optimal when contestants have asymmetric ability and are risk averse (see Sisak 2009 for a review of the theoretical literature); and penalties are useful for motivating further effort by top-tier contestants (see, for example, Moldovanu and Sela 2001, Nalebuff and Stiglitz 1983). See also Ehrenberg and Bognanno (1990), Eriksson (1999), and Harbring and Irlenbusch (2003), among others, for empirical tests of these claims.

adding competitors also makes individual contestants less likely to win, which risks diluting their incentives to exert effort in improving their performance. These basic arguments have been shown to apply both in winner-takes-all payoffs as well as in cases in which payoffs are more continuous, with multiple prizes and payoffs that increase more continuously with performance (e.g., Konrad 2007, 2009; Moldovanu et al. 2007; Moldovanu and Sela 2001).[5] This has led a number of scholars to argue that restricting the number of contestants improves contest outcomes (Che and Gale 2003, Fullerton and McAfee 1999, Nalebuff and Stiglitz 1983, Taylor 1995). The few recent empirical papers on contests in settings like sales compensations (Casas-Arce and Martínez-Jerez 2009) and test-taking (Garcia and Tor 2009) have provided some evidence of an effort-reducing impact of increased numbers of contestants. Our first prediction simply follows this basic view in the established literature.

Hypothesis 1 (Incentive Effect). *Increasing the number of competitors in an innovation contest will cause all competitors to reduce their effort, thus causing the entire distribution of performance outcomes to shift down.*

### 2.2. Innovation Contests as a Search Process

Whereas works in economics have treated different types of contests—from those concerning top managers to procurement to innovation—with the same incentive-based theoretical toolkit, more recent work, notably within the innovation and product-development literature, has taken steps to address explicitly the special character of innovation problems. This body of work places particular emphasis on innovation as a process of problem solving—or a "search" for solutions—that is subject to false steps, experimentation, serendipity, and uncertainty (e.g., Loch et al. 2001, Sommer and Loch 2004).[6] Progress might potentially be made along multiple paths and trajectories across a wide and imperfectly understood technological frontier. Therefore, stimulating innovation should involve not just incentives but also broad searching.

Because the search view of innovation shifts the focus from how any one competitor performs to how the best competitor, the winner of the contest, performs, a greater concern may thus be to design contests that increase the likelihood of at least

one extreme outcome rather than high outcomes for a large cross section of competitors (Dahan and Mendelson 2001, Girotra et al. 2010, Terwiesch and Loch 2004, Terwiesch and Ulrich 2009). Formally, if innovation attempts are independent across competitors, we may think of competitors as providing a set of random draws from some underlying distribution of possible quality of outcomes (Dahan and Mendelson 2001). If adding competitors implies adding independent solution approaches, then this would lead to a greater chance of uncovering an extreme outcome.[7]

Terwiesch and Xu (2008), in bringing this perspective alongside the formal modeling of incentives in the study of innovation contests, point out a tension between stochastic parallel path effects and incentives, particularly when the focus is on the winning performance in an innovation contest. Although they examine several institutional arrangements and contest design details, we emphasize their basic insight about the fundamental trade-off between incentives and parallel path effects as the driver of our second hypothesis.

Hypothesis 2 (Parallel Path Effect). *The negative incentive effect of increasing the number of competitors in an innovation contest will be of a smaller magnitude on the maximum performance as compared to the entire distribution of performance outcomes.*

Hypotheses 1 and 2, taken together, imply that the incentive effect will be particularly apparent on a "random" point of the distribution of outcomes, whereas the tension between the incentive effect and the parallel path effect will be more evident on the best or highest outcome. All contestants would react negatively to increased competitive pressure. However, additional "draws," as represented by additional competitors, will increase the expected maximum draw by one of the competitors. Therefore, increased rivalry has a smaller impact on the maximum performance and, as explained below, under certain circumstances might also be beneficial. It should also be mentioned that beyond simply demonstrating the distinct response of the maximum performance to added competitors, it is also crucial to gauge the *magnitudes* of the shifting of the maximum outcomes relative to the shifting in the entire distribution of outcomes. The magnitudes will tell us how important it is to consider both sets of effects when designing a contest.

---

[5] Further, the inherently public nature of contests, which often play out among individuals in socialized contexts, has led sociologists to conjecture that noncash prizes such as status and social comparison might play a role in contests, with a reduction of effort with high levels of competition (Bothner et al. 2007, Garcia and Tor 2009).

[6] This notion of innovation is a longstanding idea in the innovation literature. See, for example, Abernathy and Rosenbloom (1969), Dosi (1982), Nelson and Winter (1982), and Simon and Newell (1962).

---

[7] Consistent with the importance of searching in innovation, experimental evidence produced by Girotra et al. (2010) shows that when groups are organized in a way that leads to a higher number of ideas being generated (within a group), the best ideas are of higher quality.

## 2.3. The Moderating Effect of Uncertainty

Uncertainty is a key feature of the process of developing novel solutions to problems—in other words, of innovating (Abernathy and Rosenbloom 1969, Dosi 1982, Nelson and Winter 1982). It shapes innovation and surrounding strategic interactions in a number of ways.

A view of how uncertainty may affect contest outcomes is through heterogeneous abilities or valuations by contestants (and asymmetric uncertainty regarding them; see, for example, Konrad 2009, Terwiesch and Xu 2008). This sort of uncertainty effectively translates into uncertainty about the likelihood of any one competitor winning a contest (Konrad and Kovenock 2010, Terwiesch and Xu 2008).

Other scholars have suggested that uncertainty and its effects can be determined by focusing on the nature of a particular problem and of the knowledge required to solve it. In particular, innovation can be seen as the recombination of different sets of knowledge and ideas, thus leading to "recombinant uncertainty" (Fleming 2001, Katila 2002, Nelson and Winter 1982, Schumpeter 1943, Taylor and Greve 2006, Weitzman 1998). The greater the set of knowledge components or domains involved in addressing an innovation problem, the higher the expected uncertainty or variability of the outcomes (Fleming 2001, Taylor and Greve 2006). For example, Kavadias and Sommer (2009), in a model of problem-solving performance, consider cross-functional problems, defined as those requiring knowledge from different areas. They provide simulation results that these problems are more likely to be solved when the diversity of the solvers is fully exploited. Interestingly, even if the competitors solving a problem were identical (in contrast to the discussion in the previous paragraph), the effect of a problem being uncertain would similarly translate into uncertainty regarding the likelihood of any one competitor winning the contest.

This view of uncertainty is akin to a view of uncertain "searching" along a frontier of different paths or trajectories to improve upon existing solutions (March 1991, Sahal 1983, Silverberg et al. 1988). For challenging problems, there may be multiple fundamental approaches with varying levels of feasibility and ultimate potential. Thus, not only may competitors' ability to solve a problem differ (and the competitors not realize it) and the problem solutions have inherently high variability, but it may also not even be clear what sort of basic approach should be taken to the problem, how many possible approaches there are, and the return to pursuing any given approach. This is corroborated by Jeppesen and Lakhani's (2010) finding that the likelihood of problem-solving success for InnoCentive R&D tournaments increased with greater technical distance between the problem domain and

the solvers' own field of expertise. For the purposes of considering the issue of adding competitors in an innovation contest, this translates into uncertainty regarding the likelihood of any one competitor winning a contest.

We imply that in a contest, uncertainty in innovation often translates into uncertainty regarding precisely which competitor will achieve the best/extreme outcome (and how effective the solution of any one competitor will turn out to be). This then is the basis for the "parallel path" effect and implies that added uncertainty should simply amplify the parallel path effect: greater uncertainty increases how much adding competitors affects the maximum outcome relative to the expected distribution. Thus, the third hypothesis used to guide our empirical investigation is as follows.

HYPOTHESIS 3. *Increasing the number of competitors attempting to solve a more uncertain innovation problem will amplify the parallel path effect by having a positive impact on the maximum performance.*

It is important to note that the effect of greater uncertainty on incentives (and possible interactions with parallel paths) is a far subtler question, without a clear general prediction. Several factors are implicated here, including the shape of the knowledge distribution, number of competitors, skill levels of competitors, degree and scope of uncertainty, and so forth.[8] As an intuition, consider that if the eventual winner is from near the top of the true knowledge distribution, adding uncertainty might foster a belief in a more "level playing field" than actually exists. On one hand, this might lead eventual winners to underestimate the probability of winning and shade their level of effort downward. On the other hand, creating a perception of closer rivalry could stimulate extra effort in leaders who might otherwise "rest on their laurels." What can be argued, however, is that there should likely be some moderating effect on the incentive effect. Insofar as the moderating effect on the incentive effect could plausibly be negative, it is unclear whether added uncertainty, in the event competitors are added, should necessarily increase the net benefits to the extreme value. These issues have been

---

[8] For example, if we model uncertainty through a Gumbel distribution (see, for example, Terwiesch and Xu 2008), it can be shown that the (negative) relationship between an individual's choice of level of effort and the number of competitors is affected by the degree of uncertainty, as expressed by the scale parameter of the distribution, in a nonmonotonic way, depending on the number of competitors, the particular skill level or "draw" of a given competitor, and the scale parameter itself. See also List et al. (2010) for a study of the effect of the "slope" of the density of the random component on the competition–outcome relationship in contests.

considered in the theoretical literature only partially, and we will explore them in the empirical tests.[9]

# 3. Data and Methods

We now turn to testing empirically the hypotheses discussed above. This is challenging because an ideal empirical setting should satisfy a number of nontrivial requirements. One requirement would be the availability of precise measures of innovation outcomes. Regarding the impact of different levels of competition and uncertainty, observable measures of competitive pressure as well as metrics that distinguish problems in terms of uncertainty would be needed, as well as exogenous variations in these two characteristics. Finally, to distinguish more clearly between the effect of competitive pressure and uncertainty on the stochastic and effort components of the innovation outcomes, an ideal empirical setting would include information on the whole distribution of outcomes rather than, for example, only on the maximum (winning) performance.

The strictness of these requirements is witnessed by the absence of systematic empirical analyses of the impact of competitive pressure on performance in contests. On one hand, some of the available studies rely on random changes in the number of competitors, but mostly in lab experimental settings with hypothetical scenarios, thus lacking generalizability (see, for example, Garcia and Tor 2009). On the other hand, studies based on natural settings, such as Bothner et al. (2007), do not rely on exogenous variations.

The context and data we describe below allow for a rare possibility to rely on a quasi-experimental setting in a natural environment that is characterized by the availability of empirical measures (over the whole distribution of outcome), appropriate identification, and external validity. In addition to the quantitative data, our analysis is informed by interviews conducted with TopCoder executives and community members during the course of the study to understand the dynamics of the contest platform and various motivations that drive participation and performance. In what follows, we describe the data in detail and discuss the estimation approaches. Section 4 below presents our findings.

## 3.1. TopCoder Software Contests

The data that we analyze were provided by TopCoder. Established in 2001, TopCoder creates outsourced software solutions for IT-intensive organizations by encouraging independent programmers from around the world to compete in a regular stream of software-development contests. TopCoder's value proposition to its clients is that it can harness the value of large numbers of programmers and let the competition determine the best solutions without risking either a wrong hire or an incorrect solution. Over the years, TopCoder has served such clients as AOL, Best Buy, Eli Lilly, ESPN, GEICO, and Lending Tree, and TopCoder contestants have had the opportunity to win cash prizes, obtain third-party assessments of their skills, and signal their talent in a global competition through participation in thousands of contests. In 2009 alone, more than 11,122 programmers from around the world competed in 1,425 software-development contests for 47 clients.

TopCoder works with its clients to identify software needs that it converts into contests for its community of programmers. Contests target specific programming tasks like conceptualization, specification, architecture, component design, component development, assembly, and testing. Each contest submission is evaluated by a peer-review panel of three expert members, assessed by automatic test suites that check for accuracy and computation speed, or both. Winners are awarded predetermined cash awards (range: \$450–\$1,300 per contest) for their contributions, and the performance of all participants is converted into a continually updated rating for each contest category. Of the more than 250,000 programmers from around the world who have signed up as members, well over 40,000 have obtained ratings.[10]

Members are recruited through active outreach to college campuses worldwide and through joint sponsorship of programming competitions and events with high-profile technology firms. They are encouraged to participate and demonstrate their skills through weekly to biweekly *algorithm programming contests* in which participants compete against each other to solve three software-development problems in 75 minutes. The solutions to these problems are automatically scored via a large test suite that has been custom-tailored to each problem. Participation and performance data from the algorithm programming competitions provide the test-bed for analyzing our hypotheses.

---

[9] Rosen (1988) discusses different risk attitudes by competitors according to their relative ability. Mukherjee and Hogart (2010) propose a statistical model of how the relationship between one contestant's relative ability and her probability of winning depends on the overall number of competitors and the numbers of assigned rewards. Riis (2010) analyzes theoretically how different reward schemes affect the incentives of contestants of different ability. Bothner et al. (2007) study empirically, with data on NASCAR races, the risk-taking behavior of drivers according to their relative ranking and position, and Brown (2008) studies how motivations of "ordinary" professional golfers change when a superstar (e.g., Tiger Woods) participates in a tournament.

[10] Further details about the TopCoder context can be found in Lakhani et al. (2010).

**3.1.1. "Algorithm" Problems.** TopCoder relies on dedicated internal staff and outside consultants to design the software challenges used in the algorithm contests. A central concern for designers is to create problems that members will find both interesting and demanding and at the same time that allow TopCoder to discern between mediocre, average, and outstanding programmers. Mike Lydon (2010), chief technology officer for TopCoder and the principal designer of the algorithm contests' frameworks, explained:

> Algorithm problems test participants' ability to take an abstract problem statement and convert it into a working software program that will meet the requirements of the challenge. This requires creativity in developing solutions that rely on a broad knowledge of various algorithmic approaches, and the application of mathematical thinking in a severely time-limited context. While these problems are synthetic, the skills we assess and reward are directly applicable to diverse and demanding domains like computational biology and genomics, biomedicine, aerospace engineering, image processing, financial fraud detection, graphical rendering, and text mining, amongst many others.

Our interviews with TopCoder problem designers revealed that they have to create challenges that have well-defined outcomes so that automated test suites can be used to assess performance. An example is given by the following: "Find the most popular person in a social network of differing ethnicities in the least amount of computation time." Performance can be assessed automatically, whereas the potential approaches to solve the problem can vary. Thus, the preceding problem requires knowledge of both graph theory and string parsing to develop an effective solution. TopCoder problem designers, in their attempt to create challenges that test both ability and knowledge of a variety of algorithmic approaches, explicitly consider a variety of relevant knowledge domains that could be designed into a problem. Table 1 provides a listing of the knowledge categories used by TopCoder. Competing solvers simply access the problem statement and do not know how many or which knowledge domains the designer has designated for a particular problem.

Once a problem has been developed, TopCoder designers create an automated test suite to check for algorithmic accuracy. The test suites consist of hundreds of test cases containing obvious and nonobvious edge conditions that a programmer must meet to create the right solution to a problem. The problem designer and an experienced quality assurance engineer then simulate the test conditions by trying to solve the problems themselves within the 75-minute time constraint. Based on their experience with the problems, they assign a final maximum points value to each problem.

**Table 1    Knowledge Domains Underlying TopCoder Problems**

| Knowledge category | No. of problems |
| --- | --- |
| Encryption/Compression | 19 |
| Advanced math | 63 |
| Greedy | 84 |
| Sorting | 99 |
| Recursion | 117 |
| Geometry | 119 |
| String parsing | 128 |
| Simple search, iteration | 148 |
| Graph theory | 151 |
| Simulation | 157 |
| Search | 170 |
| String manipulation | 192 |
| Math | 202 |
| Simple math | 213 |
| Dynamic programming | 245 |
| Brute force | 251 |

*Note.* The number of problems associated with different problem types exceeds the count of problems in the population because about half of the problems are tagged as belonging to multiple categories.

**3.1.2.    Algorithm Competition Format.** Algorithm contests are held at different times and on different days of the week to accommodate TopCoder's global membership. Contest dates and times are advertised in advance to all registered members of TopCoder through personalized e-mails and on the company's website. Competitions occur in two broad divisions, I and II, based on prior skill ratings in previous algorithm contests. Division I consists of participants who rank above a predetermined rating score; Division II includes newcomers and those who rank below the Division I threshold score.

On the day of a contest, members are given a three-hour window in which to register. Five minutes before the start of the contest, registration is closed and the typically hundreds of entrants in any given contest are divided into groups, termed virtual "rooms," of not more than 20 competitors. TopCoder chose the virtual room format to accommodate large numbers of competitors, typically several hundred, in the contest without making it so intimidating and large that competitors would be discouraged. Another reason for creating virtual rooms of not more than 20 coders was to allocate prize money across the wider pool of participants. Each virtual room receives the same three problems in the division, but direct competition largely takes place within a single room. This is because rank within an individual room determines cash prizes, if any, as well as public recognition for winning. Because prizes are divided among different subsets of direct competitors by virtual room, there

**Figure 1    Illustration of the Structure of Weekly Events ("Rounds")**

| | Round 38 (Tue., 1/23/2001, 11:15 A.M. EST) | | | Round 39 (Fri., 2/2/2001, 3:00 P.M. EST) | | | Round 40 (Wed., 2/14/2001, 1:00 P.M. EST) | | |
|---|---|---|---|---|---|---|---|---|---|
| | Problem 231 | Problem 232 | Problem 233 | Problem 234 | Problem 235 | Problem 236 | Problem 237 | Problem 238 | Problem 239 |
| **Room 1** | ▓ | ▓ | ▓ | | | | | | |
| **Room 2** | ▓ | ▓ | ▓ | | | | | | |
| ... | ▓ | ▓ | ▓ | | | | | | |
| **Room 1** | | | | ▓ | ▓ | ▓ | | | |
| **Room 2** | | | | ▓ | ▓ | ▓ | | | |
| **Room 3** | | | | ▓ | ▓ | ▓ | | | |
| **Room 4** | | | | ▓ | ▓ | ▓ | | | |
| ... | | | | ▓ | ▓ | ▓ | | | |
| **Room 1** | | | | | | | ▓ | ▓ | ▓ |
| **Room 2** | | | | | | | ▓ | ▓ | ▓ |
| **Room 3** | | | | | | | ▓ | ▓ | ▓ |
| ... | | | | | | | ▓ | ▓ | ▓ |

*Notes.* This figure is an illustration of the structure of rounds at TopCoder. For example, round 39 was run on Friday, February 2, 2001, starting at 3 P.M. EST. Competitors were divided into different "virtual" rooms, and in each room the same three problems (234, 235, and 236) were assigned.

might be on the order of one to two dozen winners among several hundred entrants. As an example, Figure 1 illustrates the algorithm contest arrangement for three contests.

In the early years, from 2001 to 2003, TopCoder experimented with a range of assignment procedures to the rooms, including an "Ironman"-style assignment procedure where participants were rank-ordered by rating and then sequentially placed in a room up to capacity. Members reacted negatively to this approach, and the company converged to a simple random assignment from 2004.

Contests consist of two distinct phases: 75 minutes of programming followed by 15 minutes of solution testing. In the programming phase, participants write and submit code for each of the three problems. Each problem is assigned an amount of points visible at the start of the contest: typical values include 250, 500, and 1,000. As soon as a participant opens the problem (i.e., gets the full problem statement), the available points for a successful submission start to decline based on the amount of time between problem opening and submission of code. Hence, the faster the programmer finishes the submission, the greater the number of points available, subject to automated testing at the end. If participants open all three problems at the same time, all three will have the total number of points declining.

Competitors within individual rooms are also provided with rich information about each other and the unfolding of the competition in the room. Included in a "heads-up" display in which participants complete their code is the full list of the competitors in the room (those who have logged in following the registration period), which is color-coded to facilitate quick assessment of their skill ratings. Figure 2 presents

what competitors see. Because there are 20 or fewer competitors in a room, this information is easily navigable. The display also reveals who has submitted solutions, to enable the progression of the contest to be observed in real time. The ability to observe the submission of solutions by competitors gives participants an idea of whether they are ahead or behind in the competition.

Final scores for each participant are determined in the testing phase by automatically compiling the software code for each problem and subjecting it to automated test cases to determine the accuracy of the solution. During the testing phase, within each virtual room, participants have the right to examine any other competitor's code and submit a test case they believe would cause their competitor to fail. If the challenge test case is successful, the challenger receives 50 additional points and the challenged participant loses all points for that problem. The test case is then made part of the full, automated test suite for all participants. Challengers risk losing 25 points if they are unsuccessful in disqualifying their opponents. Performance over all of the test cases is summed and the time taken to submit the answer converted into an objective final public score and ranking of each participant's algorithm code–writing skills. Post-testing, the problem performance score and ranking of each participant within the room and in the competition are publicly released.

**3.1.3. Motivations to Participate in Algorithm Contests.** A central concern in innovation contests is the motivation of participants.[11] A chief lever

---

[11] By virtue of the contests lasting a fixed 75 minutes, the effort exerted is the level of cognitive effort rather than, say, a discretionary level of working hours or capital investment.

**Figure 2**    **Typical Public Profile of a TopCoder Competitor**

*Source.* Reproduced with permission from TopCoder. http://www.topcoder.com/tc?module=MemberProfile&cr=7442498.
*Note.* This is the typical public profile of a TopCoder competitor. It shows the skills ratings, earnings, and placement in various contests.

available to elicit participation in the contests is the structure and form of prizes. As noted earlier (in §2.1), the literature has examined both "winner-takes-all" and more continuous prize structures. The TopCoder environment in general and the algorithm contests in particular provide discrete payoffs for winners as well as more continuous payoffs across competitors. Winning cash is the most conspicuous motivation to participate in TopCoder. Between 2001 and 2010, TopCoder disbursed more than $1M in cash prizes for the algorithm contests alone. Beyond direct cash, there is a wide range of motivators that are more "continuous," whereby higher ranking outcomes generate higher payoffs. The public nature of rankings and ratings is crucial. Placing high in an individual contest or achieving a high rating through sustained success are nonpecuniary sources of satisfaction that can also directly translate into career opportunities. High-profile firms like Intel, Facebook, Google, and Microsoft, for example, both sponsor the algorithm contests and encourage some prospective employees

to obtain a TopCoder rating to be considered for programming positions. To many participants, the ratings are also a sort of status symbol. Members have their own profile pages that track performance in every contest and provide a ratings measure and distribution on TopCoder (Figure 2). Dips and rises in performance and rankings after each contest are publicly discussed on the TopCoder community message boards. Our interviews revealed that members, especially those in the higher-performing brackets, took it very personally if they did not come out on top in a competition. This point also surfaces what appears to be an intrinsic desire to compete in many members. Lydon (2010) notes, "Regardless of cash prizes, winning in the rooms and in the overall competition is everything to our top members." Thus, those who do not rank first still may receive some "prize" related to their relative position. There is, however, a major discontinuity in the reputation effect in classifying first as opposed to any other position.

These various motivators beyond just cash incentives are consistent with a number of papers that

**Table 2    Variable Definitions**

| Variable | Definition |
|---|---|
| (1) *Score* | The final score awarded to a given solution to a problem |
| (2) *No. Competitors* | Number of competitors directly competing with one another in a room |
| (3) *Average Score* | Total number of points awarded to competitors in a given room for a given problem, divided by *No. Competitors* |
| (4) *Maximum Score* | Highest or winning score within a room |
| (5) *Skill Rating* | Numerical evaluation of a competitor's skill, based on history of performance |
| (6) *Average Skill Rating* | Total *Skill Rating* in a room, divided by *No. Competitors* |
| (7) *Variance Skill Rating* | Standard deviation (second moment) of *Skill Rating* in a room |
| (8) *Skewness Skill Rating* | Skew (third moment) of *Skill Rating* in a room |
| (9) *Maximum Skill Rating* | Highest *Skill Rating* of all competitors in a room |
| (10) *No. Domains* | Count of the number of canonical problem/solution types that are part of the problem |

have remarked on the importance of sociological and behavioral motivators of various kinds in contests (Altmann et al. 2008, Konrad 2009, Moldovanu et al. 2007). These more continuous sources of performance-based payoffs appear to be rather important in at least this context. TopCoder executives noted that they observe little difference in performance whether a cash prize is offered or not, particularly now that the contest platform has grown and is internationally known by software developers (only about a third of algorithm contests have cash prizes).

### 3.2.    Data and Variables

TopCoder executives granted us access to the full database records of their, roughly weekly, algorithm contests between 2001 and 2007. Our analysis focuses on the elite Division I, in which ratings were more reliable and individual solvers tend to compete more regularly than individuals did in Division II. The sample covers 645 problems. Our empirical analysis focuses on the variation across rooms, the distinct

groups of direct competitors that compete on each problem. There are 9,661 room-problem contest observations. We first describe our outcome variables and then the key explanatory variables. Descriptions of all of the variables used in our analysis are provided in Table 2, and descriptive statistics and correlations are in Table 3.

**3.2.1.    Measuring Problem-Solving Performance.** We measure innovation performance outcomes in terms of the final score assigned by TopCoder's automated evaluation system to a given solution to a given problem, which we denote as *Score*. The *Score* per problem is based on the initial preset points allocation, which declines steadily once a competitor opens the problem during the contest up to the point of submission to the evaluation test suite. The faster a competitor codes, the higher the score, contingent on passing all challenges and system tests. Particularly relevant, given our research questions, are the average scores (*Average Score*) and maximum scores (*Maximum Score*) attained in a given room for a given problem. We also consider two additional measures of individual outcomes. Recall that the final score is the result of not just an automated set of performance tests and a barrage of test scenarios; it is further adjusted if competitors find weaknesses in the solutions of others. To assure that the final score is, in fact, a good representation of the merits of a solution rather than just representative of, say, strategic effects or a tit-for-tat challenge, we also ran our analyses using the initial submission score and a dichotomous variable that distinguishes submissions considered incorrect (value of 0) and not incorrect (value of 1).

**3.2.2.    Number of Competitors.** The main explanatory variable is *No. Competitors*: that is, the number of direct competitors facing each other in the same virtual room. For the regularly scheduled algorithm competitions, this number ranges between 10 and 20, with 99% of our sample between 15 and 20; hence, the variation that we examine is up to a 33% increase from 15 to 20 competitors. The drivers of this variation are given both by the actions of the contest

**Table 3    Descriptive Statistics and Correlations**

| Variable | Mean | Std. dev. | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| (1) *Score* | 300.1 | 300.6 | | | | | | | | | |
| (2) *No. Competitors* | 18.5 | 1.2 | −0.04 | | | | | | | | |
| (3) *Average Score* | 283.4 | 129.6 | 0.44 | −0.09 | | | | | | | |
| (4) *Maximum Score* | 313.7 | 212.7 | 0.2 | −0.02 | 0.39 | | | | | | |
| (5) *Skill Rating* | 1,689.2 | 412.1 | 0.58 | −0.05 | 0.17 | 0.12 | | | | | |
| (6) *Average Skill Rating* | 1,751.6 | 239.2 | 0.15 | −0.05 | 0.29 | 0.17 | 0.26 | | | | |
| (7) *Variance Skill Rating* | 422.9 | 164.8 | 0.09 | −0.02 | 0.21 | 0.13 | 0.15 | 0.50 | | | |
| (8) *Skewness Skill Rating* | 0.9 | 0.5 | −0.05 | 0.07 | −0.13 | −0.01 | −0.13 | −0.22 | 0.00 | | |
| (9) *Maximum Skill Rating* | 2,366.4 | 438.5 | 0.11 | 0.00 | 0.27 | 0.16 | 0.20 | 0.74 | 0.84 | −0.06 | |
| (10) *No. Domains* | 1.8 | 0.8 | −0.04 | 0.05 | −0.09 | −0.06 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 |

sponsor and by the participants. Our in-depth interviews with TopCoder participants and executives on the assignment process provide us with evidence that this variation is exogenous.

TopCoder attempts to fill each room to 20 contestants. However, in practice, participants do not arrive in groups of 20, thus creating a simple "indivisibility" problem, which will inherently create a situation in which there must be a difference of at least one across different rooms. In addition, some noise created by the room-assignment algorithm would typically generate several rooms ranging, say, from 15 to 18 participants. The other main driver of variation in the number of competitors is given by the "no-shows"; i.e., individuals who signed up and were assigned to a room but failed to checkin and participate in the contest. No-shows know neither the identities of their competitors nor the nature of the problem before deciding not to show up. Nor is their presence shown on the heads-up display in their rooms; they are effectively absent and invisible. We cannot directly observe the decision to not show up, but there are strong indirect indications of such decisions occurring. First, TopCoder managers and participants see this as a "fact of life" on the platform. We also speculate that we should see more no-shows on weekdays if it is simply harder to plan and predict one's availability. Consistent with this view, we found that the average number of participants in rooms (while keeping total participation constant) was lower on weekdays than on weekends.

**3.2.3. Level of Uncertainty.** We are also interested in how the relationship between innovation performance and number of competitors might be affected by uncertainty (§2.3). As noted by Sommer et al. (2009, p. 125), because established empirical measures of uncertainty are not readily available, researchers have to rely on the empirical context for their derivation.[12] This measure thus requires special attention and care to motivate and interpret. Discussions and interviews with TopCoder managers led us to focus on a particular source of uncertainty that appeared as the most salient: the number of problem domains on which a given solution draws. As relates to algorithm contests, TopCoder managers have long been sensitive to the need to make the problems continually interesting and challenging in order to maintain a high degree of participation. Apart from randomly mixing who appears in a given room of competitors, TopCoder's problem designers also deliberately tune and adjust the degree of uncertainty in competition outcomes. The attention to problem design

has led TopCoder to keep records of the nature of problems according to 16 problem domains (Table 1). Roughly half of the problems included in competitions are single-domain problems; that is, they are classified as belonging to just one of these 16 categories. In conforming to a given problem type, these single-domain problems have canonical solution approaches. Although they remain nontrivial, a dominant approach or template can be used to develop the solution to the problem. Anecdotal accounts from competitors strongly corroborate this contention of TopCoder managers and problem designers. The competitors suggest that approaches to these problems can often be somewhat standardized, and even possibly "routinized," at least to some extent.

Take, for example, the Bridge Building problem posted on May 18, 2006. This problem required the participants to calculate the maximum number of playing cards that could be stacked in an overlapping manner so that a bridge could be built of a certain length "*d*" over the edge of a table. TopCoder classified this as a "simple math" problem. The solution required knowledge of the basic harmonic series, and of the 264 submissions received for this problem, 82% were correct. This indicates a high degree of understanding of the problem and the requisite knowledge required to solve it. TopCoder problem designers and executives suggested that the somewhat standardized approaches used for single-domain problems were far less likely in instances in which problems drew from multiple domains. Multiple domain problems often do not just "add" two sorts of problems together such that rote solutions might still be viable. It is in combining canonical problems to produce multidomain problems that the problem designers attempt to inject greater uncertainty into performance outcomes.

A sports strategy problem posted on July 26, 2006, is exemplar of the multidomain problems. In this case, participants had to calculate the probability that players in a three-on-three basketball game reached the ideal scoring position and executed a successful basket or pass while accounting for their rivals' potential interference. Internally, TopCoder classified this as a problem straddling the knowledge domains of geometry, graph theory, and math. Overall, 338 individuals actually opened the problem with only 66 submitting solutions and 47 passing the system tests. This is an indication that the problem posed a significant challenge to even the most elite TopCoder developers.[13]

---

[12] Sommer et al. (2009), for example, relied on survey-based self-reports by managers on a Likert scale to operationalize and quantify (unforeseeable) uncertainty.

[13] A postcontest synopsis by one participant provides insight into the uncertainty faced by the competitors: "Picture yourself as an average Division 1 coder. You have just quickly finnishedy[sic] quickly the easy problem and think that there's enough time left to take the medium slow. The 50 extra points contribute to this

These considerations of TopCoder problem designers and competitors are echoed and supported by research on "recombinant" problem solving and innovation, according to which the presence of multiple knowledge domains should produce higher uncertainty and risk in the innovation process (Fleming 2001, Kavadias and Sommer 2009, Taylor and Greve 2006). This is also reminiscent of March (1991), where the exploration of multiple projects or paths leads to higher uncertainty (and to a higher likelihood of extreme outcomes). Thus, we use the number of knowledge domains from which a given problem design draws (*No. Domains*) as a measure that relates to uncertainty in both the problem approach and the eventual winner.

Beyond the earlier theoretical and context-based arguments, we conducted several tests to confirm that numbers of knowledge categories served as a meaningful indication of uncertainty.[14] We first examined whether greater uncertainty (or at least variation) in the problem leads to greater variation in score outcomes. We found that the maximum likelihood estimate of the relationship between score variance and number of problem categories (controlling for the precontest ratings of participant and for time) is positive and significant ($0.52$; $p < 0.01$), indicating a greater variance in outcomes when a problem crosses multiple knowledge categories, which is consistent with greater uncertainty. We also verified the implication that it is more difficult to predict the winner in multidomain problems—in particular, that it is less likely that the winner is the top-ranked contestant in a room. We found that the probability of the top competitor within a room (based on precontest rating) winning declines by 7% for each additional knowledge domain in a given problem.[15]

### 3.3. Estimation Approach and Control Variables

To understand how variations in *No. Competitors* affect the distribution of performance outcomes in a room

and, subsequently, how the level of uncertainty moderates this relationship, we estimate versions of the following model:

$$Y_{ij} = \alpha + \beta(\text{No. Competitors})_i + \gamma X_i + \delta_j + \varepsilon_{ij}, \quad (1)$$

and the "extended" model that also considers the role of uncertainty, as measured by the number of knowledge domains, is expressed by adding an interaction term as in the following model:

$$Y_{ij} = \alpha + \beta_1(\text{No. Competitors})_i + \beta_2(\text{No. Competitors})_i$$
$$\cdot (\text{No. Domains})_j + \delta_j + \varepsilon_{ij}. \quad (2)$$

The outcome variable $Y$, as discussed above, will be given by both *Average Score* and *Maximum Score*, and the unit of observation is a room $i$ for a given problem $j$ (in a given round).

Our greatest concern is that the coefficient estimates, especially on *No. Competitors*, might be biased by spurious correlations associated with other possible determinants of performance.[16] In principle, any number of factors might influence performance outcomes: whether a particular round had money prizes, the size of the prize(s), whether a given round received corporate sponsorship, how well-known TopCoder was at that time, how a given round corresponded to the calendar year or hiring cycles, and so forth. These factors, for example, may affect the frequency of no-shows mentioned above. Further, features of the problem, such as the maximum theoretically attainable number of points, might also have a direct influence on performance. Thus, we need to control for a wide range of variables.

We address these issues by controlling for all differences across rounds, and for time and differences across problems, by adding problem fixed effects in the regressions—represented by $\delta_j$ in models (1) and (2) above. In doing so, we identify the relationship of interest out of the differences across rooms for a given problem.[17] This radically simplifies the estimation problem. Consequently, the task of the control variables—the matrix $X_i$ above—is to account for differences across rooms. The rooms themselves are identical; however, what varies across rooms is *who* is in them. Therefore, what remains to be controlled is the composition of the individuals in a room because, again, the presence of different types of competitors might affect the decision to participate actively or not.

---

impression. After reading the problem statement, you write down some numbers and mathematical expressions, maybe think about a dynamic programming approach, but nothing convinces you. After 15 minutes of doing this, you are mad at yourself and take a look at the division summary: Nobody has submitted! Not even one of the many coders that have several hundred rating points more than you. . . . Solving this problem required imagination and either faith or a good proof. The strange thing in this case is that almost everybody solved it differently." Summary written by TopCoder member Soul-Net, available at http://www.topcoder.com/tc?module=Static&d1=match_editorials&d2=srm313.

[14] We thank our anonymous reviewers for suggesting these tests.

[15] The estimated coefficient on the number of domains is not statistically significant at conventional levels but is robust to including or dropping control variables.

[16] Given that these are one-shot competitions and participants do not have details about the problems or direct competitors before a competition starts, we can reasonably rule out reverse causality.

[17] Note that the coefficient on the "direct" term for *No. Domains* cannot be independently estimated because we are controlling for problem fixed effects; therefore, this term is not present in model (2).

TopCoder provides an excellent measure of the skills of all participants based on their historical performance. Every competitor is evaluated and rated after each contest using the long-established "Elo" system used to evaluate, rate, and rank chess grandmasters (van der Maas and Wagenmakers 2005). This system assesses skills based on the performance of a competitor relative to everyone else working on the same problem and is dynamically updated after each contest. The *Skill Rating* is used to rank all participants on the TopCoder platform, and we use it to as a control for compositional differences across competition rooms striving to solve the same problem.

Therefore, the intuition for the empirical approach is that we estimate how varying the number of competitors across rooms for a given problem affects the distribution of outcomes, controlling for differences in the distribution of skills across rooms.[18]

The regression technique will therefore be given by linear panel models with problem fixed effects. Note that problem fixed effects control for differences across not only individual problems but also different rounds and time. We will also estimate quantile regression models (using a weighted absolute deviation algorithm; see Koenker and Basset 1978, and Koenker and Hallock 2001). As argued in more detail below, quantile regressions help to separate the impact of changes in competition and uncertainty on various parts of the distribution of outcomes and, in particular, to contrast these changes with changes in the observed maximum score in a room. In doing so, these analyses will allow for a clearer distinction between incentive effects and parallel path effects for different levels of competitive pressure and problem uncertainty. Moreover, the empirical exploration of any differential impact of competition on different parts of the distribution of outcomes might offer insights to theory because little is known on this particular point.

## 4. Results

Our results are reported in three subsections. In §4.1, we assess the baseline model and its robustness in its simplest form. In §4.2, we report results of quantile regressions to show how the wider distribution of problem-solving performance outcomes changed with the number of competitors, and we contrast this with

how the maximum (winning) problem-solving performance outcome was affected by varying numbers of competitors. Finally, in §4.3, we report how varying levels of uncertainty moderate the relationships tested earlier.

### 4.1. The Baseline Model

We begin by estimating the baseline model by simply relating *Average Score* to *No. Competitors*. The regressions, therefore, estimate how the average score in a room changed on average with varying numbers of competitors, or the average incentive effect. Following Hypothesis 1, we should expect a negative coefficient on *No. Competitors*. Results are presented in Table 4.

Column 4-1 reports estimates from regressing *Average Score* on *No. Competitors* with problem fixed effects.[19] The coefficient estimate is negative and highly significant. To ensure that differences across rooms are not biasing the estimated coefficient, what remains is to control for compositional differences across rooms. In column 4-2 the simplest measure of differences in skills across rooms, *Average Skill Rating*, is added. This changes the magnitude of the coefficient on *No. Competitors*, but the coefficient remains negative ($-5.08$) and highly significantly different from zero. As an assessment of the effectiveness of controls for skill across rooms, we examine alternative specifications. First, we allow for the possibility that the effect of *Average Skill* may enter nonlinearly, as in column 4-3 where different average skills levels in different rooms are broken into 20 individual dummies at each five-percentile increment of the variable. This does not change the results. To control for the distribution of skills within a given room more fully, column 4-4 adds the variance, skewness, and maximum of *Skill Rating*.[20] This also does not change the results, with the average score in this estimate still declining by about five points with each added competitor ($-4.63$; $p = 0.01$).

To corroborate the meaningfulness of the *Score* measure (which represents the final score conferred on a given solution), we assay several alternative problem-solving performance variables as dependent variables. We find the same patterns, whether using the

---

[18] Whereas *Score* may vary appreciably from problem to problem, we find that the distribution of the estimated residuals from a regression of *Score* on problem fixed effects yields a much smoother, single-peaked distribution. Moreover, analyses not reported here (but available upon request) show that the *No. Competitors* and *Skill Rating* do not vary systematically from problem to problem, suggesting that these variables are not strongly correlated with particular problems.

[19] The *F*-test for the overall model fit is significant at $p = 0.01$ for all models. Standard error estimates are robust to autocorrelation and heteroskedasticity.

[20] Using the log of the skills rating yields similar estimates of the coefficient on *No. Competitors*. We also assessed the robustness of results with a completely different approach in which we estimated how individual competitors' performance varied from round to round, controlling for individual competitor fixed effects, a series of covariates for round and problem covariates (using event time of day as an instrumental variable). This approach produced almost identical point estimates of the average effect of added competitors, but with lower statistical significance.

**Table 4    Baseline Fixed-Effect Regressions of Performance Outcomes** (*Average Score*) **on Numbers of Competitors** (*No. Competitors*)

| Dependent variable: | *Average Score* | | | | Alternative dependent variables | |
|---|---|---|---|---|---|---|
| | 4-1 | 4-2 | 4-3 | 4-4 | 4-5 | 4-6 |
| Explanatory variables | Problem fixed effects | Control for avg. room skills | Flexible, nonlinear control for avg. skills | Control skills distribution (preferred) | Avg. submission score | Fraction of submission score "not incorrect" |
| *No. Competitors* | −9.24*** (0.84) | −5.08*** (0.73) | −5.01*** (0.73) | −4.63*** (0.72) | −4.68*** (0.72) | −0.0036** (0.0020) |
| *Skill Rating* distribution | | | | | | |
| *Average* | | 0.18*** (0.00) | | 0.14*** (0.01) | 0.17*** (0.01) | 0.0001*** (0.0000) |
| *Average* (dummies for different bands) | | | Yes | | | |
| *Variance* | | | | 0.02** (0.01) | 0.04*** (0.01) | −0.0000*** (0.0000) |
| *Skewness* | | | | −22.16*** (1.53) | −26.03*** (1.58) | −0.0250*** (0.0020) |
| *Maximum* | | | | 0.02*** (0.01) | 0.00 (0.01) | 0.0000 (0.0000) |
| Problem fixed effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Adjusted *R*-squared | 0.60 | 0.70 | 0.70 | 0.72 | 0.74 | 0.73 |

*Notes.* Robust standard errors are in parentheses. Number of observations is 9,661 room-problems.
   *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

initial submission score (before potential challenges— see §3.2.1), as in column 4-5, or an indicator for simply not being incorrect (column 4-6).[21]

We also perform a number of additional robustness checks, which are reported in Table 5. One concern might be that the pattern of a roughly five-point average decline with each added competitor might be limited to the range of variation in the number of competitors that we observe; the bulk of the data in our sample constitutes rooms comprising between 15 and 20 competitors and focuses on weekly online contests. To assess this possibility, we supplement our sample with data from ad hoc contests held by TopCoder, which tended to qualify competitors for an annual in-person, sponsored event called the TopCoder Open. Although these contests are different from those in our main sample, they essentially follow the same rules of the game, use similar Web-based facilities, and involve the same sort of algorithmic problems. Crucially, the number of competitors in these instances ranges from less than 15 to more than 20. Estimates from regressions using these data are almost identical to those reported in Table 4. Column 5-1 reports results from regressions based on the main data and the extended out-of-sample data.

To ensure that the sample estimates do not confound or obscure cases in which there were no monetary prizes for a contest, we also run the regressions on the subsample of contests without prizes.

We find no differences in the estimates as reported in column 5-2. This finding is consistent with the opinion of TopCoder executives and direct observation of these competitions. Finally, we ensure that there are no major differences in the estimates across different rounds or events based on unobserved factors such as whether a contest received sponsorship. Although we do not observe these details, we do observe total attendance at a given event (the sum of competitors across rooms). Estimating the models on subsamples of widely attended versus sparsely attended events should thus provide some indication of the robustness of the results across different sorts of events. We do so by putting all observations from contests with below-median participation (for a given year) in one sample and those with above-median participation in another. The results in both cases appear similar to those for the entire sample as reported in columns 5-3 and 5-4.[22]

---

[21] We estimate a linear probability model in this latter case. Binary models (logit or probit) convey the same results.

[22] An additional point that we felt important to document relates to changes over time. Although individual problem-level fixed effects control for time trends per se, we were also interested in whether the relationship between performance and number of competitors might itself change through time. We ran the model on early (pre-2004) and late (post-2004) subsamples and found that the coefficient estimate on *No. Competitors* in both subsamples is negative and significant, which is consistent with earlier results. However, we noted that the pre-2004 coefficient estimate is statistically different from the earlier estimates, being −9.66 (s.e. = 1.73). A closer year-by-year examination reveals that this is driven by a lone, aberrant correlation in the first year of the sample, 2002, at −13.4 (s.e. = 9.0). The estimate for 2002 has multiple possible explanations: a potentially different profile of early participants at TopCoder; the different

**Table 5  Baseline Fixed-Effect Regressions on Subsamples and Out-of-Sample Data**

| Dependent variable: | *Average Score* | | | |
|---|---|---|---|---|
| | 5-1 | 5-2 | 5-3 | 5-4 |
| Explanatory variables | Extended range of *N* (out of sample) | No prize | "Big" rounds | "Small" rounds |
| *No. Competitors* | −5.39*** | −5.74*** | −4.21*** | −5.56*** |
| | (0.68) | (0.88) | (0.87) | (1.26) |
| *Skill Rating* distribution | | | | |
| *Average* | 0.16*** | 0.15*** | 0.14*** | 0.14*** |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| *Variance* | −0.01 | 0.02** | 0.03*** | 0.00 |
| | (0.01) | (0.01) | (0.01) | (0.02) |
| *Skewness* | −9.66*** | −24.93*** | −20.46*** | −25.75*** |
| | (1.51) | (1.83) | (1.85) | (2.73) |
| *Maximum* | 0.02*** | 0.02*** | 0.01** | 0.01*** |
| | (0.00) | (0.01) | (0.01) | (0.01) |
| Problem fixed effects | Yes | Yes | Yes | Yes |
| Adjusted *R*-squared | 0.72 | 0.72 | 0.73 | 0.67 |
| Observations | 13,156 | 7,219 | 4,831 | 4,830 |

*Note.* Robust standard errors are in parentheses.
    *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

### 4.2. Number of Competitors, Distribution of Outcomes, and Extreme-Value Performance

Having observed a general negative shift in performance outcomes with added competitors, we now examine the effects on the overall distribution of outcomes. Results are presented in Table 6. The analysis begins by documenting how the wider distribution of outcomes (beyond just the average) shifted in response to added competitors. We present quantile regression results (from a model analogous to expression (1) in §3.3) for the 25th, 50th (median), 75th, and 90th percentiles of the distribution of outcomes in columns 6-1–6-4.[23] The coefficients for each quantile are estimated to be negative, suggesting the negative incentive effect is general, which is consistent with Hypothesis 1. The upper tail of performance shifts downward more than the rest of the distribution does, as seen in the greater magnitude of the estimated coefficients for upper quantiles. This might be so for a number of reasons. It could be that leading competitors are responding more (negatively) to competition; however, it might simply reflect that the points scale is bounded on the low end at zero or other features of

the points system. Therefore, we cannot offer a definitive interpretation based on these data and analyses and leave that for future theory and empirical investigations.

Having demonstrated that the negative response to added competitors is general across the distribution of outcomes and is particularly negative for the higher quantiles, we now examine how adding competitors affects the maximum score. The maximum score should reflect not just shifting incentives and effort but also the stochastic parallel path effect, which can create additional upside for the maximum score attained. In this context (i.e., when the distribution shifts downward), we should observe a less negative response of the maximum score to increasing *No. Competitors* than was seen in the overall downward shift in the distribution. In column 6-5, estimates for which the dependent variable is the maximum score in a room are reported. The maximum score decreases by only −0.88 with each additional competitor, and this estimate is not significantly different from zero. Consistent with Hypothesis 2, the maximum score effectively shifts upward *in relation to the distribution of outcomes*.

Figure 3 presents the results of quantile regressions (at 5% increments) graphically and contrasts these with the response of the maximum. Thus, in summary, we find support for Hypotheses 1 and 2. The downward shift in the distribution of outcomes is consistent with a negative incentive effect. The upward shift in the maximum in relation to this distribution is consistent with a parallel path effect

---

level of socialization in the early, small TopCoder community; and other possible differences in the early TopCoder. Given that the inclusion of these data does not meaningfully change results and we cannot account for the difference, we simply continue to include these early data in our estimates.

[23] An alternative way to demonstrate similar patterns is to simply regress both the mean and variance of room scores on *No. Competitors*. This approach produces results in line with those of the quantile regressions.

**Table 6      Quantile Regressions**

| Outcome variable (quantile): | *Score* | | | | |
|---|---|---|---|---|---|
| | 6-1 | 6-2 | 6-3 | 6-4 | 6-5 |
| Explanatory variables | q25 | q50 | q75 | q90 | Maximum |
| *No. Competitors* | −1.74*** | −1.68*** | −5.08*** | −9.39*** | −0.88 |
| | (0.56) | (0.55) | (1.18) | (1.82) | (1.52) |
| *Skill Rating* distribution | | | | | |
|   *Average* | 0.09*** | 0.12*** | 0.23*** | 0.33*** | 0.11*** |
| | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) |
|   *Variance* | −0.02** | 0.00 | 0.07*** | 0.19*** | 0.04** |
| | (0.01) | (0.01) | (0.01) | (0.02) | (0.02) |
|   *Skewness* | −15.01*** | −21.13*** | −35.52*** | −19.86*** | 6.25* |
| | (1.14) | (1.14) | (2.47) | (3.75) | (3.32) |
|   *Maximum* | 0.00 | 0.00 | −0.01 | −0.02 | 0.02** |
| | (0.00) | (0.00) | (0.01) | (0.01) | (0.01) |
| Problem fixed effects | Yes | Yes | Yes | Yes | Yes |
| Constant | −178.59*** | −55.33*** | 117.45*** | 333.26*** | 372.10*** |
| | (0.54) | (0.53) | (1.15) | (1.77) | (32.36) |
| Observations | 162,561 | 162,561 | 162,561 | 162,561 | 9,661 |

*Notes.* Models in columns 6-1–6-4 are estimated with weighted least absolute deviations, with standard errors following Koenker and Bassett (1978, 1982). Model 6-5 is estimated with ordinary least squares and robust standard errors.

    *, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

coexisting with the incentive effect because the maximum score should benefit from greater numbers of draws.
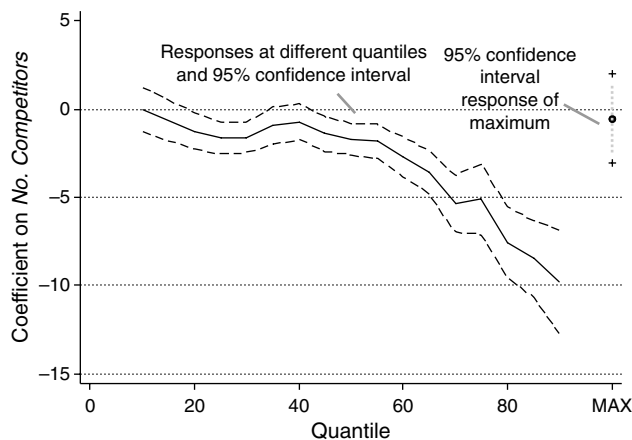
### 4.3. The Effect of Uncertainty

We now examine how varying levels of uncertainty (as associated with the number of problem domains) affect the earlier relationships (Hypothesis 3). We expect that greater uncertainty would increase the

**Figure 3      Change Across the Distribution of Performance Outcomes with Added Competitors**



*Notes.* Each point on the solid line measures the relationship between performance and *No. Competitors* at the respective quantile, controlling for a fixed effect for the particular problem being solved and controlling for the distribution of skills of individuals within a given room. The dotted lines represent the 95% confidence interval. The response of maximum score is shown at the 100% position (precisely the maximum); the dot is the coefficient estimate, with 95th percent confidence intervals shown above and below.

magnitude of the parallel path effect, meaning that the maximum score should shift upward from the distribution of outcomes to a greater degree when there is higher uncertainty. The results to follow confirm this point.

The models estimated here are essentially the same as the earlier models, but they add an interaction between the proxy for uncertainty and the number of competitors (see expression (2), §3.3). Results are presented in Table 7. The preferred models for *Average Score* (Table 4, column 4-4) and *Maximum Score* (Table 6, column 6-5) are also included in this table for comparison, in columns 7-1 and 7-3.

We do not have a strong prior belief regarding the exact functional form through which the measure of uncertainty, *No. Domains*, maps to levels of uncertainty, only that uncertainty should increase with *No. Domains*. Therefore, we assess several functional forms of how this could enter into the interaction with *No. Competitors*. We report estimates from models for which we interact *No. Competitors* with a binary indicator for multiple domains (i.e., *No. Domains* > 1), thus distinguishing between single- and multidomain problems.[24] We estimate the model with this indicator interacted with *No. Competitors* for both the *Average Score* and the *Maximum Score* (columns 7-2 and 7-4). The coefficient estimate on the interaction term for the *Maximum Score* (column 7-4) is large and positive

---

[24] Models based on different specifications, such as entering *No. Domains* as a linear or quadratic interaction (rather than a binary variable), lead to similar but less statistically powerful results.

**Table 7  Fixed-Effect Regressions on the Moderating Effect Uncertainty** (*No. Domains*)

| Dependent variable: | Average Score | | Maximum Score | |
|---|---|---|---|---|
| | 7-1 (4-4) | 7-2 | 7-3 (6-5) | 7-4 |
| Explanatory variables | No interactions | Interaction with multidomain problems | No interactions | Interaction with multidomain |
| *No. Competitors* | −4.63*** | −8.07*** | −0.88 | −5.46** |
| | (0.72) | (1.16) | (1.52) | (2.44) |
| *No. Competitors* $\times I\{Multiple\ Domains\}$ | | 5.75*** | | 7.65** |
| | | (1.46) | | (3.10) |
| *Skill Rating* distribution | | | | |
| Average | 0.14*** | 0.14*** | 0.11*** | 0.11*** |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Variance | 0.02** | 0.02** | 0.04** | 0.04* |
| | (0.01) | (0.01) | (0.02) | (0.02) |
| Skewness | −22.16*** | −22.04*** | 6.25* | 6.40* |
| | (1.53) | (1.53) | (3.32) | (3.32) |
| Maximum | 0.02*** | 0.02*** | 0.02** | 0.02** |
| | (0.01) | (0.01) | (0.01) | (0.01) |
| Problem fixed effects | Yes | Yes | Yes | Yes |
| Adjusted *R*-squared | 0.72 | 0.72 | 0.53 | 0.53 |

*Notes.* Robust standard errors are in parentheses. Number of observations is 9,661 room-problems.
*, **, and *** indicate statistical significance at the 10%, 5%, and 1% levels, respectively.

(7.65). The coefficient on the interaction term for *Average Score* (column 7-2) is also estimated to be positive and significant but smaller (5.75). The greater positive effect of adding uncertainty on the maximum score is consistent with the presence of a parallel path effect acting on the maximum score (i.e., the change in the average score should only reflect changing incentives

and not parallel paths). Whereas the higher uncertainty of multidomain problems moderates the negative (incentive) response to the average score, the net effect remains negative (i.e., −8.07 + 5.75). By contrast, high uncertainty in the multidomain problems leads to a net positive effect (i.e., −5.46 + 7.65).

To show more explicitly how added competitors reshaped distributions of outcomes with varying levels of uncertainty, we plot results of the 10th, 25th, 50th, 75th, and 90th quantile regressions and the maximum score linear regression results in Figure 4. Again we see a general downward shift in the distribution of outcomes with added competitors. However, the effect on single-domain/low-uncertainty problems is more negative for all quantiles. We also see that the maximum score moves upward in relation to the distribution with added competitors. Hence, we find support for Hypothesis 3, and we further note that uncertainty in the problem being solved also shifts the incentive response. Table 8 provides a summary of our main findings, links to the existing literature, and the resultant implications.

**Figure 4  Change in Response Due to Added Competitors for Single- and Multidomain Problems**
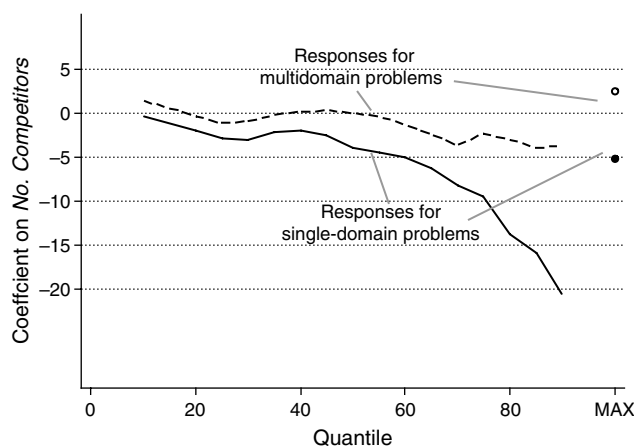


*Notes.* Each point measures the relationship between performance and *No. Competitors* at the respective quantile, controlling for a fixed effect for the particular problem being solved and controlling for the distribution of skills of individuals within a given room. The solid black line relates to single-domain problems; the dashed black line relates to multidomain problems. The response of maximum score is shown at the 100% position (precisely the maximum), with the single-domain problems shown as the solid black dot and multidomain problems shown as the white dot.

## 5.  Summary and Conclusions

Why do innovation contest organizers typically invite and encourage widespread entry? Most economic models of tournaments suggest that widespread entry should diminish contest performance by reducing incentives to exert effort for *all* competitors (Che and Gale 2003, Fullerton and McAfee 1999, Taylor 1995). One explanation, suggested by the work of innovation scholars is that added competitors may lead to more

**Table 8    Summary of Findings and Contributions**

| Investigation | Findings | Related literature | Implications |
|---|---|---|---|
| Incentive effects | Adding competitors leads to a downward shift in the entire distribution of outcomes. | Bothner et al. (2007), Che and Gale (2003), Fullerton and McAfee (1999), Garcia and Tor (2009), Taylor (1995) | Generalized free entry in all types of innovation contests is not to be recommended. |
| | Downward shift appears to impact higher percentile of performance distribution. | | Further theorizing and empirical investigation is required to see if incentive effects impact participants with varying skills and abilities in a different way. |
| Parallel path effects | The effect of adding competitors on the observed maximum score is less negative (not different from zero) than the distribution of outcomes—consistent with the presence of a "parallel path" effect. | Dahan and Mendelson (2001), Girotra et al. (2010), Terwiesch and Ulrich (2009) | The order statistic effect is observed and confirmed with innovation contests. |
| | The magnitude of parallel path effect is of the same order as that of incentive effects. | | Both effects exist in innovation tournaments, and negative incentive effects countervail positive parallel path effects. |
| Role of problem uncertainty | Increasing uncertainty moderates the competition–performance relationship across the entire distribution of outcomes. | Terwiesch and Xu (2008) | Higher uncertainty problems dampen incentive effects. Competitors do not respond to rivalry as much when faced with highly uncertain problems. |
| | Higher uncertainty mitigates and potentially reverses the negative impact of added competitors on the maximum performance. | | Free entry should be encouraged only in contests for which problems are highly uncertain. |
| | The moderating effect of uncertainty is stronger for higher percentiles of the outcome distribution. | | Higher uncertainty may lead to more effort by competitors who are near the top of the performance distribution. |

independent experimentation or parallel path effects, increasing contest performance (particularly the top score) as more competitors are added (Abernathy and Rosenbloom 1969, Dahan and Mendelson 2001, Nelson 1961). These arguments go to the very heart of a key question in the design of innovation contests: how many competitors to let in.

Given the increasing importance of contests to elicit innovation, systematic empirical evidence is required to document and quantify the presence of, workings of, and interplay among these effects; we sought to estimate each of these distinct effects, assess their relative importance, and understand under which conditions one effect might dominate the other. Analyzing detailed microdata from 9,661 contests, we find the following patterns:

1. *Negative Incentive Effect Across the Entire Distribution of Performance Outcomes*. Our findings provide empirical confirmation that adding competitors shifts expected outcomes downward. This result provides support that a negative "incentive effect" is at work across the full distribution of outcomes. We also find that the downward shift in our performance measure is greater for higher percentiles of the distribution.

2. *Coexistence of Parallel Paths Produces Effects of Similar Magnitude*. Although the distribution of outcomes may generally shift downward as competitors are added, on account of incentive effects, the maximum maximum or top score shifts upward in relation to this distribution. Thus, the maximum score responds more positively (less negatively) to added competitors than does the distribution of performance outcomes. Thus, we detect the presence of parallel path effects coexisting alongside incentive effects. Adding competitors thus generates the "upside" potential of achieving an extreme outcome. Although abundant theory presumes this effect, we contribute to a nascent literature (see, for example, Girotra et al. 2010) that quantifies it. More importantly, that incentive and parallel path effects were both large and of comparable magnitudes implies that neither should be ignored when modeling or designing contests.

3. *Moderating Effect of the Level of Uncertainty and the Nature of the Problem Being Solved*. We are able to observe "degrees" of uncertainty in our context by recording the nature of the problem. Single-domain problems are more certain in how they would be solved and who would solve them best, whereas multidomain problems are less certain in these regards. We find that higher uncertainty not only increases the (positive) parallel path effect of adding competitors but also reduces the (negative) incentive effect. Thus, the moderating effect of uncertainty is very strong—so much so that the net effect of adding competitors

on the top score is positive in the cases of multidomain (high-uncertainty) problems. The effect on the top score remained negative in single-domain (low-uncertainty) problems. Hence the underlying problem uncertainty is a crucial parameter in the design of innovation contests.

Beyond these findings, it is noteworthy that our study provided a "natural" setting in which we were able to observe both the distribution of skills and the distribution of outcomes of competitors, across multiple groups of direct competitors, or independent contests with varying numbers of competitors for a given problem—and for hundreds of problems. These multiple "trials" per problem were then also crucial for devising our empirical approach.

Our results have implications for managers organizing innovation contests. Managers need to be aware that contests set in motion opposing effects in response to the number of competitors allowed to participate. The practitioner literature has mostly celebrated the virtues of open entry (e.g., Lindegaard 2010, Tapscott and Williams 2006); however, realizing that, by definition, most participants *lose* and that increasing competition decreases individual incentives (even at the highest quantiles of the performance distribution) should cause managers to have realistic expectations of both the benefits and drawbacks of "open" innovation contests. An additional implication is that to attract and retain solvers who otherwise may not be "winners," managers might do well to explicitly create ancillary benefits of participation such as learning, career signaling, and community identification.[25] Managers might also want to consider changing the design of contests such that participation information is revealed strategically, perhaps after the contest, so that the incentive effect does not dominate the parallel path effect. This is at least consistent with our empirical finding that greater uncertainty diminished the strength of (negative) incentive effects on performance.

Our findings about the role of uncertainty in mediating between the incentive and parallel path effects also highlight the essential role managers must play in selecting and/or designing the innovation problems to be resolved through contests. In particular, managers need to design contests such that the free entry criterion is reserved for problems with a high degree of uncertainty. Alternatively framed, we might reserve wide, open innovation contests, intended to attract a large number and variety of competitors, as appropriate when we have exhausted conventional approaches and the contest institution is exploited

to bring a diversity of approaches—and a potential upside from widespread experimentation. Certainly, historical examples such as those described in the introduction of this paper would appear to suggest this use of open contests as not just incentive mechanisms but, rather, as ways to attract diverse perspectives (Jeppesen and Lakhani 2010) and to sort and select individuals with peculiar preferences (Boudreau and Lakhani 2011).

A few limitations of our study should also be mentioned. The first is the nature of the problems being solved in our empirical context. Although these were challenging problems and demanded considerable cognitive effort by elite software developers, they were explicitly devised by professional designers with the goal of creating challenges for competitors. It was, in fact, this very characteristic that produced observable measures to characterize the problems. We might expect many innovation problems to be more uncertain than were those problems observed here with respect to the technical approaches to their solutions. Therefore, what we emphasize in our findings is not the relevance of the absolute level of uncertainty of the problems studied but the ability to differentiate patterns over a varying range of uncertainty.

We might also imagine problems that are more or less responsive to effort and incentives compared to those studied here. The problems studied here were considerably smaller in scale than are typical industrial or scientific innovation problems. Their regularity, small scale, and recurrence were essential to our ability to study general patterns, but we might expect a range of additional factors to play a role in larger-scale problems that are drawn out over longer periods and perhaps embedded in more complex team or organizational dynamics. However, it should be noted that contests are increasingly seen as a regular and ongoing platform for innovation (Boudreau and Hagiu 2009) rather than just an innovation approach reserved for special and sporadic ad hoc events, as was perhaps historically the case. From this perspective, the modular problem solving that occurs at regular intervals might be seen as less extraordinary.

Another potential limitation is that in these contests, we do not observe varying work hours, capital investments, or other discretionary levels of investment. Although the general patterns observed conform to theory, the results more directly reflect contests in which we observe the behavior of individual people who compete rather than a context in which, say, firms decide investment levels more generally. The central point and emphasis of the paper is that incentive effects (whatever their nature and origin) appear to coexist with the parallel path effect.

The results presented in this article suggest that neither order-statistic arguments related to parallel paths

---

[25] Boudreau and Lakhani (2009) offer examples of heterogeneous motivations used in innovation platforms.

nor game theoretic arguments related to strategic incentives should be ignored in modeling or designing innovation contests. This is not just a substantive finding in its own right but also suggests that current traditions of modeling innovation contests (i.e., modeling just one set of mechanisms without the other) may largely ignore key interactions and trade-offs. To our knowledge, only Terwiesch and Xu (2008) have begun to make progress in integrating these issues thus far.

## Acknowledgments

## References

Abernathy, W. J., R. S. Rosenbloom. 1969. Parallel strategies in development projects. *Management Sci.* **15**(10) 486–505.

Altmann, S., A. Falk, M. Wibral. 2008. Promotions and incentives: The case of multi-stage elimination tournaments. IZA Discussion Paper 3835, Institute for the Study of Labor, Bonn, Germany.

Bothner, M. S., J. Kang, T. E. Stuart. 2007. Competitive crowding and risk taking in a tournament: Evidence from NASCAR racing. *Admin. Sci. Quart.* **52**(2) 208–247.

Boudreau, K. J., A. Hagiu. 2009. Platform rules: Multi-sided platforms as regulators. A. Gawer, ed. *Platforms, Markets and Innovation*. Edward Elgar, London, 163–191.

Boudreau, K. J., K. R. Lakhani. 2009. How to manage outside innovation: Competitive markets or collaborative communities? *Sloan Management Rev.* **50**(4) 69–76.

Boudreau, K. J., K. R. Lakhani. 2011. The confederacy of software. J. Lerner, S. Stern, eds. 50*th Anniversary Volume of NBER Rate and Direction of Inventive Activity*. National Bureau of Economic Research, Cambridge, MA. Forthcoming.

Brown, J. 2008. Quitters never win: The (adverse) incentive effects of competing with superstars. Working paper, Northwestern University, Evanston, IL.

Casas-Arce, P., F. A. Martínez-Jerez. 2009. Relative performance compensation, contests, and dynamic incentives. *Management Sci.* **55**(8) 1306–1320.

Che, Y.-K., I. Gale. 2003. Optimal design of research tournaments. *Amer. Econom. Rev.* **93**(3) 646–671.

Dahan, E., H. Mendelson. 2001. An extreme-value model of concept testing. *Management Sci.* **47**(1) 102–116.

Dosi, G. 1982. Technological paradigms and technological trajectories: A suggested interpretation of the determinants and directions of technical change. *Res. Policy* **11**(3) 147–162.

Ehrenberg, R. G., M. L. Bognanno. 1990. Do tournaments have incentive effects? *J. Political Econom.* **98**(6) 1307–1324.

Eriksson, T. 1999. Executive compensation and tournament theory: Empirical tests on Danish data. *J. Labor Econom.* **17**(2) 262–280.

Fleming, L. 2001. Recombinant uncertainty in technological search. *Management Sci.* **47**(1) 117–132.

Fullerton, R. L., R. P. McAfee. 1999. Auctioning entry into tournaments. *Amer. Econom. Rev.* **107**(3) 573–605.

Garcia, S. M., A. Tor. 2009. The *N*-effect: More competitors, less competition. *Psych. Sci.* **20**(7) 871–877.

Girotra, K., C. Terwiesch, K. T. Ulrich. 2010. Idea generation and the quality of the best idea. *Management Sci.* **56**(4) 591–605.

Harbring, C., B. Irlenbusch. 2003. An experimental study on tournament design. *Labour Econom.* **10**(4) 443–464.

Harris, C., J. Vickers. 1987. Racing with uncertainty. *Rev. Econom. Stud.* **54**(1) 1–21.

Holmstrom, B. 1982. Moral hazard in teams. *Bell J. Econom.* **13**(2) 324–340.

Jeppesen, L. B., K. R. Lakhani. 2010. Marginality and problem-solving effectiveness in broadcast search. *Organ. Sci.* **21**(5) 1016–1033.

Katila, R. 2002. New product search over time: Past ideas in their prime? *Acad. Management J.* **45**(5) 995–1010.

Kavadias, S., S. C. Sommer. 2009. The effects of problem structure and team diversity on brainstorming effectiveness. *Management Sci.* **55**(1) 1899–1913.

King, R. 2000. *Brunelleschi's Dome: How a Renaissance Genius Reinvented Architecture*. Penguin, New York.

Knight, F. H. 1921. *Risk, Uncertainty and Profit*. Harper, New York.

Koenker, R., G. Basset, Jr. 1978. Regression quantiles. *Econometrica* **46**(1) 33–50.

Koenker, R., G. Bassett, Jr. 1982. Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* **50**(1) 43–61.

Koenker, R., K. F. Hallock. 2001. Quantile regression. *J. Econom. Perspect.* **15**(4) 143–156.

Konrad, K. A. 2007. Strategy in contests. WZB–Markets and Politics Working Paper SP II 2007-01, Social Science Research Center Berlin (WZB), Berlin.

Konrad, K. A. 2009. *Strategy and Dynamics in Contests*. Oxford University Press, Oxford, UK.

Konrad, K. A., D. Kovenock. 2010. Contests with stochastic abilities. *Econom. Inquiry* **48**(1) 89–103.

Lakhani, K. R., D. Garvin, E. Lonstein. 2010. TopCoder (A): Developing software through crowdsourcing. HBS Case 610-032, Harvard Business School, Boston.

Lazear, E. P., S. Rosen. 1981. Rank-order tournaments as optimum labor contracts. *J. Political Econom.* **89**(5) 841–864.

Lindegaard, S. 2010. *The Open Innovation Revolution: Essentials, Roadblocks and Leadership Skills*. Wiley, Hoboken, NJ.

List, J. A., D. van Soest, J. Stoop, H. Zhou. 2010. On the role of group size in tournaments: Theory and evidence from lab and field experiments. Working paper, University of Chicago, Chicago.

Loch, C. H., C. Terwiesch, S. Thomke. 2001. Parallel and sequential testing of design alternatives. *Management Sci.* **47**(5) 663–678.

Lydon, M. 2010. Personal Communication, February 16.

March, J. M. 1991. Exploration and exploitation in organizational learning. *Organ. Sci.* **2**(1) 71–87.

McKinsey & Company. 2009. And the winner is: Capturing the power of philanthropic prizes. Accessed October 19, 2010, http://www.mckinsey.com/clientservice/Social_Sector/our_practices/Philanthropy/Knowledge_highlights/And_the_winner_is.aspx.

Moldovanu, B., A. Sela. 2001. The optimal allocation of prizes in contests. *Amer. Econom. Rev.* **91**(3) 542–558.

Moldovanu, B., A. Sela, S. Xianwen. 2007. Contests for status. *J. Political Econom.* **115**(2) 338–363.

Mukherjee, K., R. M. Hogart. 2010. The *N*-effect: Possible effects of differential probabilities of success. *Psych. Sci.* **21**(5) 745–747.

Nalebuff, B. J., J. E. Stiglitz. 1983. Prizes and incentives: Towards a general theory of compensation and competition. *Bell J. Econom.* **14**(1) 21–43.

Nasar, J. L. 1999. *Design by Competition: Making Design Competitions Work.* Cambridge University Press, Cambridge, UK.

National Research Council. 2007. *Innovation Inducement Prizes at the National Science Foundation.* The National Academies Press, Washington, DC.

Nelson, R. R. 1961. Uncertainty, learning, and the economics of parallel research and development efforts. *Rev. Econom. Statist.* **43**(4) 351–364.

Nelson, R. R., S. G. Winter. 1982. *An Evolutionary Theory of Economic Change.* Belknap Harvard, Cambridge, MA.

Riis, C. 2010. Efficient contests. *J. Econom. Management Strategy* **19**(3) 643–665.

Rosen, S. 1988. Promotions, elections and other contests. *J. Institutional Theoret. Econom.* **144** 73–90.

Sahal, D. 1983. Technological guideposts and innovation avenues. *Res. Policy* **14**(2) 61–82.

Schumpeter, J. 1943. *Capitalism, Socialism and Democracy.* Harper, New York.

Scotchmer, S. 2004. *Innovation and Incentives.* MIT Press, Cambridge, MA.

Silverberg, G., G. Dosi, L. Orsenigo. 1988. Innovation, diversity and diffusion: A self-organisation model. *Econom. J.* **98**(393) 1032–1054.

Simon, H. A., A. Newell. 1962. Computer simulation of human thinking and problem solving. *Monographs Soc. Res. Child Behav.* **27**(2) 137–150.

Sisak, D. 2009. Multiple-prize contests: The optimal allocation of prizes. *J. Econom. Surveys* **23**(1) 82–114.

Sommer, S. C., C. H. Loch. 2004. Selectionism and learning in projects with complexity and unforeseeable uncertainty. *Management Sci.* **50**(10) 1334–1347.

Sommer, S. C., C. H. Loch, J. Dong. 2009. Managing complexity and unforeseeable uncertainty in startup companies: An empirical study. *Organ. Sci.* **20**(1) 118–133.

Tapscott, D., A. D. Williams. 2006. *Wikinomics: How Mass Collaboration Changes Everything.* Penguin, New York.

Taylor, C. R. 1995. Digging for golden carrots: An analysis of research tournaments. *Amer. Econom. Rev.* **85**(4) 872–890.

Taylor, A., H. Greve. 2006. Superman or the Fantastic Four? Knowledge combination and experience in innovative teams. *Acad. Management J.* **49**(4) 723–740.

Terwiesch, C., C. H. Loch. 2004. Collaborative prototyping and the pricing of custom-designed products. *Management Sci.* **50**(2) 145–158.

Terwiesch, C., K. Ulrich. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities.* Harvard Business School Press, Boston.

Terwiesch, C., Y. Xu. 2008. Innovation contests, open innovation, and multiagent problem solving. *Management Sci.* **54**(9) 1529–1543.

van der Maas, H. L. J., E.-J. Wagenmakers. 2005. A psychometric analysis of chess expertise. *Amer. J. Psych.* **118**(1) 29–60.

Weitzman, M. L. 1998. Recombinant growth. *Quart. J. Econom.* **113**(2) 331–360.

White House. 2010. Guidance on the use of challenges and prizes to promote open government. Memorandum for the Heads of Executive Departments and Agencies, Office of Management and Budget, Washington, DC. http://www.whitehouse.gov/omb/assets/memoranda_2010/m10-11.pdf.